Geoscience Frontiers 7 (2016) 21–31

HOSTED BY

Contents lists available at ScienceDirect

China University of Geosciences (Beijing)

Geoscience Frontiers

journal homepage: www.elsevier.com/locate/gsf

Research paper

Predictability of US tornado outbreak seasons using ENSO and northern hemisphere geopotential height variability

Kent H. Sparrow, Andrew E. Mercer*

Department of Geosciences, Mississippi State University, 108 Hilbun Hall, P. O. Box 5448, Mississippi State, MS 39762-5448, USA

ARTICLE INFO

Article history:
Received 4 August 2014
Received in revised form
20 July 2015
Accepted 31 July 2015
Available online 18 August 2015

Keywords:
Interannual variability
Support vector regression
Climate predictability
Tornado outbreaks

ABSTRACT

The predictability of dangerous atmospheric phenomena such as tornado outbreaks has generally been limited to a week or less. However, recent work has demonstrated the importance of the Rossby wavetrain phasing over the United States in establishing outbreak-favorable environments. The predictability of Rossby wavetrain phasing is strongly related to numerous climate-scale interannual variability indices, which are predictable many months in advance. To formalize the relationship between interannual variability indices and seasonal tornado outbreak frequency, indices derived from monthly mean Northern Hemisphere 500-hPa and 1000-hPa geopotential height fields and Niño 3.4 indices for ENSO phase were compared to annual tornado outbreak seasonal frequencies. Statistical models predicting seasonal outbreak frequency were established using linear (stepwise multivariate linear regression—SMLR) and nonlinear (support vector regression—SVR) statistical modeling techniques.

The stepwise methodology revealed predictors that are important in establishing outbreak-favorable environments at long lead times. Additionally, the results of the statistical modeling revealed that the nonlinear SVR technique reduced root mean square errors produced by the control SMLR technique by 28% and provided more consistent forecasts. A preliminary physical analysis revealed that years with high outbreak frequencies were associated with the presence of 500-mb troughs over the central and western US during the peak of outbreak season, while lower frequencies were consistent with ridging over the US or northwest flow over the Plains. These patterns support the results of the statistical modeling, which demonstrate the utility of geopotential height variability as a predictability measure of outbreak frequency.

1. Introduction

The U.S. typically has more than 7 major tornado outbreaks each year, resulting in substantial loss of life and property (Doswell et al., 2006). These dangerous outbreaks are typically large-scale, synoptically driven events that have reasonably good short term (up to 3 days in advance) predictability. However, no long-term tornado outbreak predictive scheme exists. Recent work by Mercer et al. (2009) and Mercer et al. (2012) recognized distinct synoptic environments in the 500-hPa geopotential height field associated with enhanced tornado outbreak potential. In particular, Mercer et al. (2012) noted distinct 500-hPa geopotential height patterns that can be associated with enhanced tornado outbreak potential (Fig. 1). These results suggested that the amplitude and phase of the Rossby wavetrain is important for outbreak formation, and that shifts in the wavetrain pattern due to interannual variability in the height fields could impact outbreak occurrence on an annual scale. Most long-term forecasting techniques utilize interannual variability patterns in the Rossby wavetrain in the midlatitudes (Shabbar et al., 2001; Durkee et al., 2008; Linkin and Nigam, 2008) and thermodynamic characteristics (e.g. sea surface temperatures—SSTs) in the tropics (Smith et al., 1998; Cook and Schaefer, 2008; Hagemeyer, 2011). These studies, as well as those described below, all demonstrate a strong relationship between the interannual variability patterns and the phasing of the Rossby wavetrain. Based on this predictability in the Rossby wave pattern, this study seeks to determine the relationship between these interannual variability indices and seasonal tornado outbreak activity, as a means of providing a long-range outbreak forecast that is currently absent from operational meteorology.

* Corresponding author.
E-mail address: mercer@hpc.msstate.edu (A.E. Mercer).
Peer-review under responsibility of China University of Geosciences (Beijing).

http://dx.doi.org/10.1016/j.gsf.2015.07.007

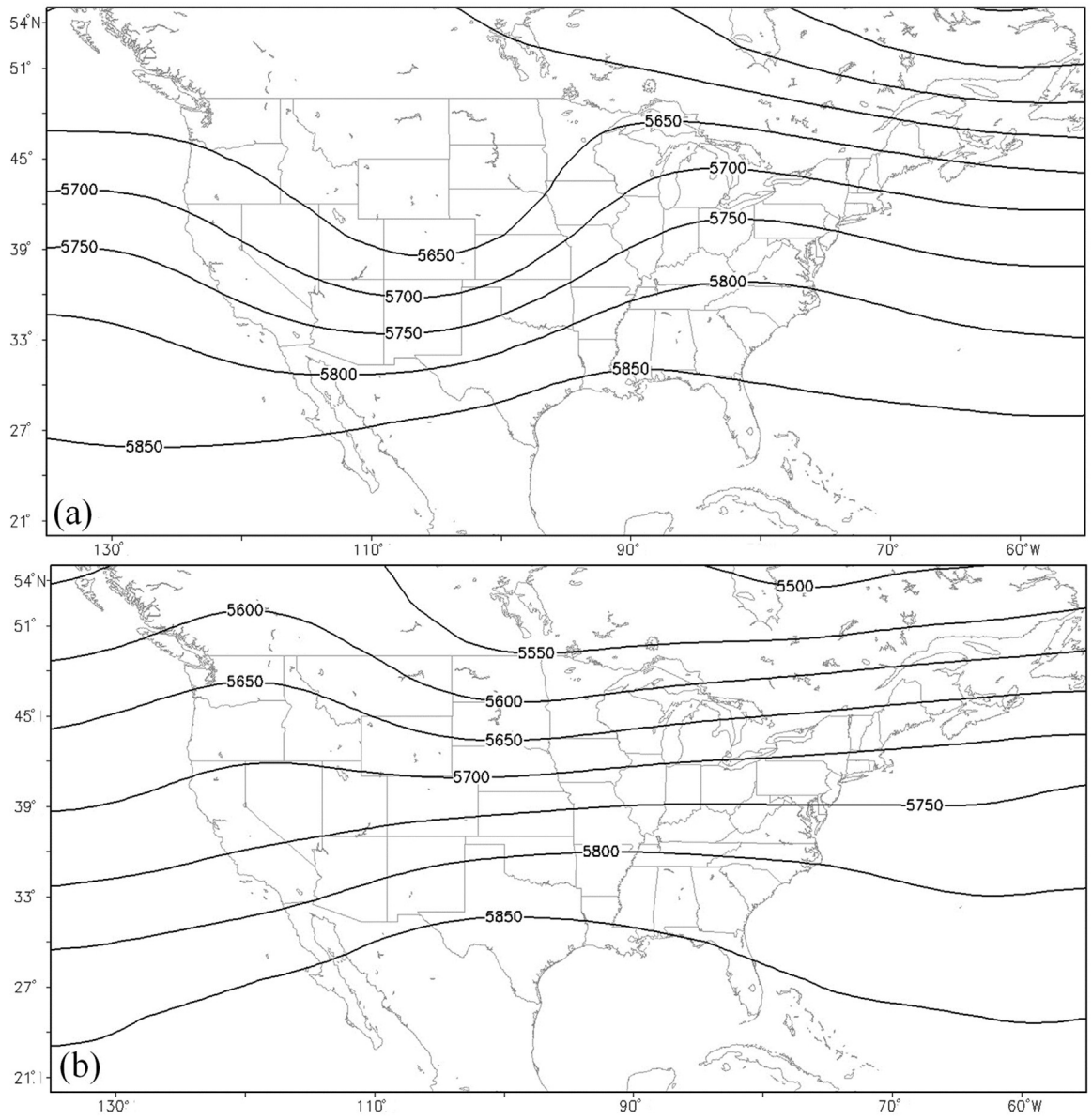

Figure 1. 500-hPa geopotential height patterns associated with (a) a tornado outbreak and (b) a primarily non-tornadic outbreak. Greater wave amplification in (a) suggests the importance of phasing in the Rossby wavetrain for tornado outbreaks. Results based on data from Mercer et al. (2012).

Wallace and Gutzler (1981) was the first to provide extensive documentation of 500-hPa height variability patterns in the Northern Hemisphere. In their work, spatial relationships between geographically distinct regions were determined by computing correlations between an individual gridpoint and all other gridpoints in their Northern Hemisphere domain. Barnston and Livezey (1987) updated the work of Wallace and Gutzler (1981) by utilizing Varimax rotated principal component analyses (RPCA – Richman, 1986) to identify the dominant modes of variability within the Northern Hemisphere 700-hPa height field. Their work identified and named many of the common patterns utilized in climate forecasting today, including the North Atlantic Oscillation (NAO), the Pacific North American oscillation (PNA), and the West Pacific oscillation (WP). This work was updated more recently by Richman and Mercer (2012), who utilized Promax RPCA on the 500-hPa geopotential height fields in the NCEP/NCAR reanalysis database (Kalnay et al., 1996), thereby relaxing the orthogonality constraint on Varimax rotation and expanding the database to include the most recent 20 years (1948–2010). Richman and Mercer (2012) found that their RPC score indices were more physically intuitive than the results found by Barnston and Livezey (1987), and the methods utilized in Richman and Mercer (2012) will be employed in the formulation of interannual variability indices for this study. In addition to the work in the mid-level geopotential height fields, Thompson and Wallace (1998) identified an oscillation in the low-level (1000-hPa) geopotential height fields that was shown to directly modify the Rossby wave amplitude, the Arctic Oscillation (AO). Additionally, recent work by Cook and Schaefer (2008) and Lee et al. (2013) revealed a relationship between tornado outbreak occurrence and the phase of the El Niño Southern Oscillation (ENSO). All of these previously established indices clearly modify the Rossby wavetrain over North America, either directly or otherwise, and as such were utilized in this study for long-term outbreak prediction.

While several studies have noted the relationship between the Rossby wavetrain and individual interannual variability indices, no

study has considered the predictability of a tornado outbreak season through the use of these indices. As such, this study will utilize 500-hPa and 1000-hPa interannual height variability indices, as well as ENSO indices, to try to predict tornado outbreak frequency the following year. A support vector regression (SVR; Cristianini and Shawe-Taylor, 2000) model (a learning machine method) will be formulated and verified using these variability indices to predict future outbreak occurrence. Note that this study does not seek the prediction of individual outbreaks, but instead will formulate a model to predict a number of outbreaks in a given season (a climate scale feature). Such predictions are likely heavily dependent upon the given year's average position of the jet stream and associated meridional amplitudes of the Rossby wavetrain, which are all shown in previous work to be heavily dependent on the considered variability indices.

Section 2 of this paper describes the datasets used and methods employed in the formulation of the SVR. Section 3 describes the model performance and verification. Section 4 provides discussion and conclusions.

2. Data and methods

2.1. Data description

To accomplish the aforementioned goals of this work, a database of tornado outbreak occurrence (with an associated outbreak definition), monthly 500-hPa and 1000-hPa Northern Hemisphere geopotential heights, and ENSO indices were required. An overview of each dataset is provided below.

i Outbreak data

Doswell et al. (2006) provided one of the first efforts in ranking tornado outbreaks, defining a tornado outbreak as 6 or more tornado reports associated with a single synoptic-scale system in a 24-h period (1200 UTC–1159 UTC). Shafer and Doswell (2010) updated this ranking system, compiling a database of outbreaks of all types from 1960 to 2013 (54 years) and ranked the relative outbreak severity using their N15 index. This ranking was based both social factors and storm reports, including variables such as the number of tornado reports, number of violent tornado reports, track length, and the outbreak's destructive potential index (DPI; Thompson and Vescio, 1998). According to Shafer and Doswell (2010), particularly high N15 index values were consistent with the strongest outbreaks that contained numerous strong to violent tornadoes. While no formal outbreak definition exists, a blended definition from Doswell et al. (2006) and Shafer and Doswell (2010) was implemented to define outbreaks in this study. In particular, an outbreak was defined as an event with 6 or more tornado reports associated with a single synoptic-scale system that also had an N15 ranking index greater than 0.75. This approach ensured that only the largest tornado outbreaks were included in the study. Interestingly, this outbreak definition yielded a statistically significant increasing trend over time of roughly 0.75 major outbreaks per decade (significant at $p = 0.008$; Fig. 2). This trend is likely affected by both anthropogenic trends in tornado reports and climate variability, though the anthropogenic effects are likely minimal since major outbreaks are rarely missed, even early in the database.

For each tornado outbreak year, tornado outbreak counts from March–December were retained. Recent work by Dixon et al. (2014) demonstrated a relative minimum in tornado activity (and thus tornado outbreak activity) in the cold season months, particularly January and February. In removing these months, the major tornado outbreak months were still included, but this approach lessened the temporal extent of the defined tornado outbreak season. For this study, the years 1960–2010 were used, with 2011 through 2013 withheld as independent testing years.

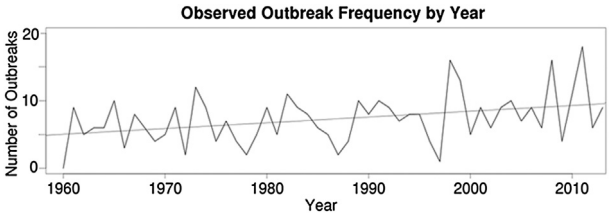

Figure 2. Total number of outbreaks by year, as defined in this study (more than 6 tornadoes and an N15 ranking index larger than 0.75). A trend line is added to show the increase in outbreak number by year currently being observed.

ii Geopotential height data

While outbreak frequencies were used as a predict and in this study, interannual variability indices were required as independent predictors of outbreak number. In particular, fields which influence the Rossby wavetrain (Mercer et al., 2012) were thought to be important for outbreak season predictability, so interannual height variability indices at 500-hPa and 1000-hPa, as well as a representation of ENSO through the Niño 3.4 index, were calculated. Following the methodology of Richman and Mercer (2012), which formulated updated interannual variability indices for Northern Hemisphere 500-hPa geopotential heights, monthly means of 1000- and 500-hPa geopotential height data from 1948 to 2013 were obtained from the NCEP/NCAR reanalysis dataset (Kalnay et al., 1996). Since variability associated with the Rossby wavetrain tends to maximize in the cold season, only fields from January and February were retained (December was tested but still showed characteristics consistent with a warm to cool season transition month).

Since the NCEP/NCAR reanalysis data are provided on a latitude-longitude grid, gridpoints will converge with increasing latitude, artificially increasing the spatial correlation between gridpoints towards the North Pole. To alleviate this issue, the height data were interpolated onto an equally-spaced hemispheric Fibonacci grid (Fig. 3; Swinbank and Purser, 2006) using a Barnes analysis (Barnes 1964).

2.2. Formulation of interannual variability indices

To update the interannual height variability indices displayed in Richman and Mercer (2012), an RPCA (Wilks, 2011) was conducted on the monthly means of 500-hPa Northern-Hemispheric geopotential heights from 1948 to 2013. This approach updates the results of Richman and Mercer through 2013 and provides additional indices and patterns for February. The RPCA methodology utilized herein followed the same approach as Richman and Mercer (2012) (e.g. S-mode analysis, Promax rotation). A scree test (Wilks, 2011) was used as a first-guess on the correct number of RPCs to retain for each month (e.g. Fig. 4), and this number was formalized following the method of the congruence coefficient test (Richman and Lamb 1985; Table 1). Both January and February tests suggested that eight 500-hPa interannual variability patterns contained the most dominant features in the monthly 500-hPa flow. The resulting RPC score indices retained small correlations (Promax rotation yields correlated RPC scores due to its obliquity) that were less than 0.2 for all indices. These small correlations were deemed insignificant for modeling purposes and all eight indices for each month were retained.

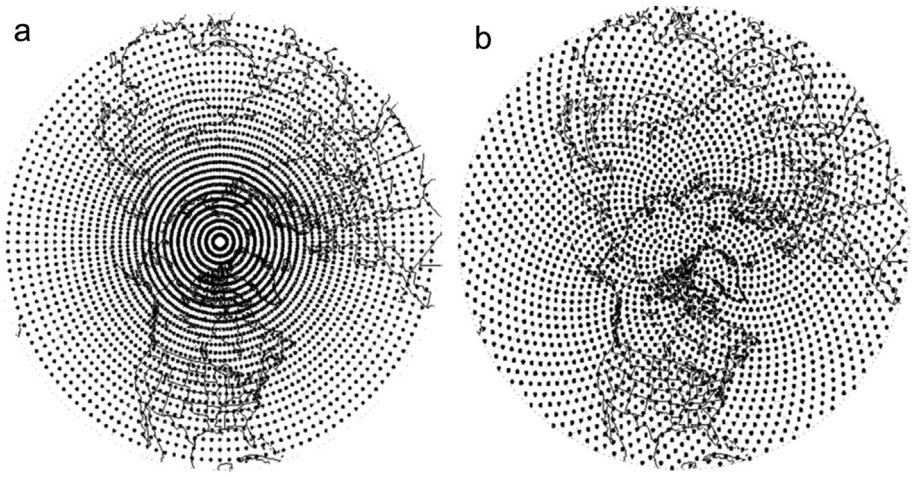

Figure 3. The NCEP/NCAR reanalysis grid (a) and the interpolated Fibonacci grid (b) (from Richman and Mercer 2012).

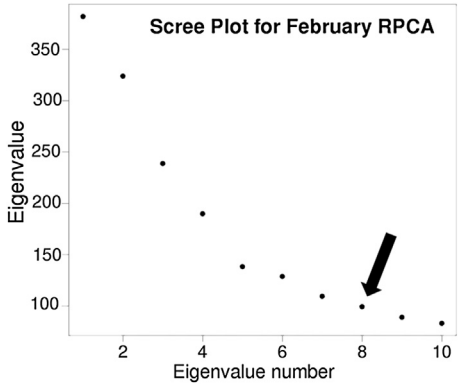

Figure 4. The eigenvalues of 500-hPa February heights. The truncation point (indicated by the arrow) is identified at the location where the slope of the eigenvalues becomes small (near zero). Similar plots were used at 1000-hPa and for January.

In addition to updating the 500-hPa height variability fields, RPCA of the 1000-hPa height fields was completed to identify the Arctic Oscillation (AO; Thompson and Wallace 1998). As was true at 500-hPa, this RPCA followed the methodology of Richman and Mercer (2012). Based on the results of the combined scree test and congruence coefficient tests for each month of 1000-hPa data, 4RPCs were retained for January and February. However, only one RPC represented the AO as portrayed by Thompson and Wallace (1998), RPC1 for both January and February. The resulting RPC loadings for both 500-hPa and 1000-hPa represented common hemispheric geopotential height variability patterns, while the RPC scores represented the corresponding phase for the given month.

The combination of the 8 RPC 500-hPa geopotential height indices, RPC1 of the 1000-hPa heights (the equivalent of the AO), and the ENSO indices for January and February made up the predictor suite used to predict seasonal numbers of tornado outbreaks. These data were used as input into statistical prediction schemes to provide forecasts of outbreak numbers by season. *Note that these indices were lagged one year behind their associated outbreak prediction* (e.g. 1959 indices were used to predict 1960 outbreak numbers).

Table 1
Example congruence (η) test for February 500-hPa geopotential heights. RPCs were truncated until all congruence values had an absolute magnitude greater than 0.81 per Mercer et al. (2012). In this example, 8 RPCs were retained.

Congruence test example			
RPC number	η	RPC number	η
1	−0.9227936	1	0.9146373
2	−0.9176560	2	−0.9176560
3	0.9451322	3	0.9451322
4	0.9174108	4	0.8114874
5	−0.8891221	5	−0.9126568
6	0.9035758	6	0.9027417
7	−0.9491890	7	−0.9523185
8	−0.7808855	8	0.8814217
9	0.8515701		

2.3. Statistical modeling methodology

i Overview of support vector regression

Support vector regression (SVR; Cristianini and Shawe-Taylor, 2000) is a nonlinear machine learning technique that has been shown in numerous studies (e.g. Mercer et al., 2008, 2009) to improve upon traditional linear modeling methods in terms of prediction. SVR falls into a special class of learning machine methods known as kernel methods, as a kernel function is the primary way by which the SVR introduces nonlinearity into the predictive system. The assumption is that the relationships between the predictor variables and the predictand are inherently nonlinear (a safe assumption in meteorological studies). To introduce nonlinearity into the system, a kernel map function φ is applied to the predictor matrix. This map function acts to project the data into a higher dimensional hyperspace where a linear fit will describe the relationship between the variables more effectively. This procedure is completed by first defining a kernel matrix:

$$\mathbf{K} = \langle \varphi(\mathbf{x}), \varphi(\mathbf{x}) \rangle_H \quad (1)$$

defined as the dot product of the kernel map function with itself (here $\mathbf{x}$ represents a vector in the original data matrix and H denotes the kernel matrix is in nonlinear hyperspace). This kernel matrix can be readily formulated from any nonlinear function. Common kernel matrix formulations include:

$$\text{Polynomial, } k\left(\mathbf{x}, \mathbf{x}^T\right) = \left(\mathbf{x}^T\mathbf{x} + 1\right)^p \quad (2)$$

$$\text{Linear, } k\left(\mathbf{x}, \mathbf{x}^T\right) = \mathbf{x}^T\mathbf{x} \quad (3)$$

$$\text{Radial Basis Function, } k\left(\mathbf{x}, \mathbf{x}^T\right) = e^{-\gamma||\mathbf{x}-\mathbf{x}^T||^2} \quad (4)$$

where γ is a measure of Gaussian spread in the radial basis function (a tunable parameter) and p represents a polynomial degree.

The formal goal of SVR is to find a function $f(\mathbf{x})$ that has falls within some tolerance ε so that:

$$\left|f(\boldsymbol{x_j}) - \boldsymbol{y_j}\right| \leq \varepsilon_j \text{ for every } \boldsymbol{x_j}$$

$$\varepsilon_j \geq 0 \text{ for every } \boldsymbol{x_j} \quad (5)$$

This dependence on this loss function ε gives this type of SVR the name ε-SVR. To formulate this function $f(\mathbf{x})$, this non-linear problem has to be solved:

$$f(\mathbf{x}) = \left\{\mathbf{x} \in \Re \rightarrow \left\langle \mathbf{w} \cdot \varphi(\mathbf{x})_{\boldsymbol{H}} + \mathbf{b} \right\rangle : \|\mathbf{w}\| \leq B\right\} \quad (6)$$

where $B > 0$ and $\mathbf{w} = \sum w_j \varphi(x_j)$. To solve for f, a nonlinear quadratic programming optimization problem must be solved:

$$\min\left(\boldsymbol{\alpha}^T \mathbf{K} \boldsymbol{\alpha} + C\boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} + b^2\right)$$

$$s.t. \ |\mathbf{K}\boldsymbol{\alpha} + b1 - \mathbf{y}| \leq \boldsymbol{\varepsilon}\prime \quad (7)$$

where C is a cost (penalty) function and $\mathbf{y}$ is a vector of y_j values (the predictand). The optimal solution for $(\boldsymbol{\alpha}^*, b^*)$ yielding a final formulation of $f(\mathbf{x})$ is:

$$f : \mathbf{x} \rightarrow \sum_j \alpha_k^* k\left(\mathbf{x}, \mathbf{x}^T\right) + b^* \quad (8)$$

where all non-zero α_k^* values are called support vectors.

ii Stepwise multivariate linear regression (SMLR) methodology

Typically, the role of SVR in modeling study is to assess its improvement over a more traditional linear modeling method (e.g. least squares linear regression). To provide a control modeling framework for comparison with the SVR model performance and conduct a variance-based forward selection of predictors, a stepwise multivariate linear regression model (SMLR) was constructed on the project dataset. SMLR is commonly utilized in regression-based studies as a control model (e.g. Ohring, 1972; Colquhoun and Riley, 1996; Billet et al., 1997), as it provides a linear construction of individual variance explained added by each predictor sequentially. Such an approach offers additional information about the contributions of individual predictors towards the predict while still retaining the relatively simple linear structure of the model. Thus, the added benefit of nonlinear modeling methods such an SVR can be assessed using SMLR as a baseline.

The stepwise forward-selection procedure was completed based on minimizing mean square error (MSE) and maximizing R^2, similar to the approach utilized in identifying significant predictors for the SVR. The predictor row at which the MSE began to increase by adding further predictors was identified as the truncation point for the stepwise, meaning all variables that reduced MSE and increased R^2 significantly were kept. Note that the stepwise was only completed on the full 51-year dataset; no cross-validation was done in the predictor selection step due to computational constraints. For this study, the predictors $\mathbf{x}$, reduced by this stepwise procedure, represent the important outbreak predictors (Table 2) for all years from 1959 to 2009, lagged one year behind their associated tornado outbreak counts ($\boldsymbol{y}$).

Table 2
Decision output for the SMLR. All tested RPC score indices which were considered significant by the stepwise analysis are displayed, along with their associated MSE and R^2 values, in order of decreasing importance according to the stepwise procedure.

Stepwise regression results for significant predictors		
Variability index	MSE	R^2
Jan. RPC 2	10.484	0.132
Jan. RPC 7	9.035	0.252
Feb. RPC 4	8.259	0.33
Feb. RPC 8	7.921	0.371
Jan. RPC 3	7.439	0.422
Feb. AO	7.268	0.447
Feb. RPC 3	7.049	0.476
Jan. RPC 1	6.895	0.499
Jan. ENSO	6.836	0.515
Feb. ENSO	6.678	0.537
Feb. RPC 6	6.322	0.573

iii Support vector regression optimization methodology

As is clear from the above formulation, several tunable parameters exist within the SVR framework. In particular, the cost function C, which modifies the importance of observations far from the hyperplane fit (higher costs increase the likelihood of overfitting the model to the support vectors), the loss function ε which can be modified to prevent overfitting the SVR, and the kernel matrix function and its associated parameters. Modifications to any of these parameters can drastically affect the outcomes in the SVR, so care must be taken when choosing a proper set of parameters.

To identify the optimal kernel-cost-loss combination, 120 different experiments were cross-validated using a bootstrap k-fold cross validation routine. Cross-validation was completed by randomly withholding 2 of the 51 years as independent testing data and computing root mean square error (RMSE) statistics on the testing set. The bootstrapping was done 1000 times and in a pairwise manner to ensure the training/testing sets were consistent between kernel-cost-loss combination experiments. In addition to testing the previously listed kernel functions (Eqs. 2–4), the following values for loss and cost functions were considered:

- ε-loss, 0.01, 0.005, 0.001.
- Cost, 1, 10, 100, 1000, 10,000.

The bootstrap cross-validation approach yielded confidence intervals of RMSE for each of the 120 experiments. Initially, any experiment whose lower RMSE limit exceeded the median RMSE of a different experiment was rejected outright. This led to the rejection of all high cost squared-polynomial combinations, as these yielded median RMSE values of roughly triple the remaining experiments (8.8 outbreaks vs. 3 outbreaks). After removing these initial poor performing combinations, the kernel-cost-loss combination with the smallest median RMSE was retained as the best predicting combination. Such an approach assured the best possible performance. Additionally, interquartile ranges (IQR) of the bootstrap replicates for the retained kernel-cost-loss combinations were determined to ensure minimal spread relative to the SMLR bootstrapped baseline IQR (2.43). This ensured that the RMSE was within the minimum confidence level of all experiments and had the smallest spread in its replicates, yielding the most consistent results, which is important in a forecasting application. The best performing combination was the radial kernel with $\varepsilon = 0.05$, $\gamma = 0.01$, and a cost of 10.

To support the results of the "optimal" combination described previously, SVR performance based on ε, cost, and kernel function

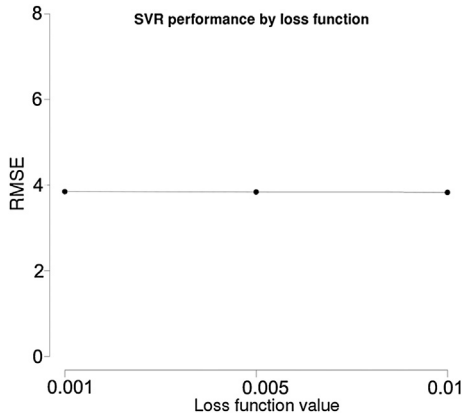

Figure 5. Performance of the SVR based on the value of ε. Results are based on all kernel-cost-loss combinations for the given value of ε.

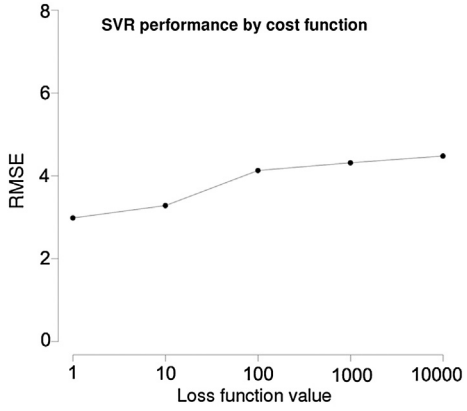

Figure 6. Same as Fig. 5, but for the cost function.

was assessed. This analysis was completed by using all tested instances of a given value (e.g. all kernel and cost experiments with a $\varepsilon = 0.01$). Such an analysis revealed several interesting characteristics of the SVR training phase. First, the SVR model seemed to be largely independent on the value of the loss function ε (Fig. 5), while the SVR performance suffered considerably with increasing cost value (Fig. 6). The poorer performance with increasing cost is likely due to overfitting the model with the higher cost values, which is an expected behavior with SVR. The lack of sensitivity of the results to ε likely means that the model converges on a solution prior to this threshold, possibly suggesting its sensitivity to overfitting, which is consistent with the cost results. Finally, model performance based on the kernel function was diagnosed (Fig. 7). It was clear that all polynomial kernel functions, with the exception of the 0.5 root kernel, behaved much worse than any radial basis function kernel, and that all radial basis function kernels performed increasingly well as the value of γ was reduced. Interestingly, the 0.5 root kernel performed similarly to the small γ radial basis function kernels, but its results were more sensitive to higher costs and thus it was not selected as the optimal kernel function. Overall, these results support the conclusion that the best kernel-cost-loss combination for this SVR was identified.

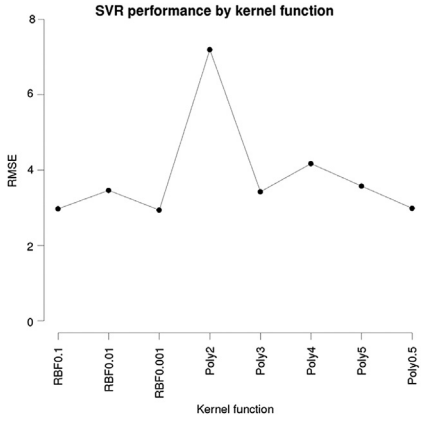

Figure 7. Same as Fig. 5, but for each kernel function.

3. Results

3.1. Statistical modeling results

Prior to the stepwise analysis, initial linear regression results using all tested predictors yielded an R^2 of 0.606, suggesting a strong relationship between the predictors and the tornado outbreak frequency. However, the stepwise analysis showed that MSE values increased after keeping 11 predictors, yielding a local minimum of MSE of 6.322 (hence an RMSE of 2.51 outbreaks per year) from the training set (Table 2). Additionally, this reduction in the larger predictor set only decreased the R^2 to 0.573, which still suggests strong predictability of tornado outbreaks a year in advance.

As stated previously, the SMLR was cross-validated using a bootstrap k-fold routine with 1000 repetitions. The median RMSE from this bootstrap cross-validation at a one-year lag was approximately 2.28 outbreaks per year, a result that was consistent with the stepwise results (Table 2). Additionally, a box and whisker plot of the RMSE bootstrap replicates (Fig. 8, left box plot) revealed minimal spread in the results, with an interquartile range of 2.43 and the whisker spread extending from zero to approximately eight outbreaks per year. The RMSE distribution was also slight platykurtotic (−0.67) and skewed towards higher values (skewness of 0.62, likely due to the lower boundary of zero in the distribution).

When assessing the SVR model error, the associated box and whisker plot RMSE median value was approximately 1.96 outbreaks per (Fig. 8, right box plot). This result only suggested roughly a 0.32

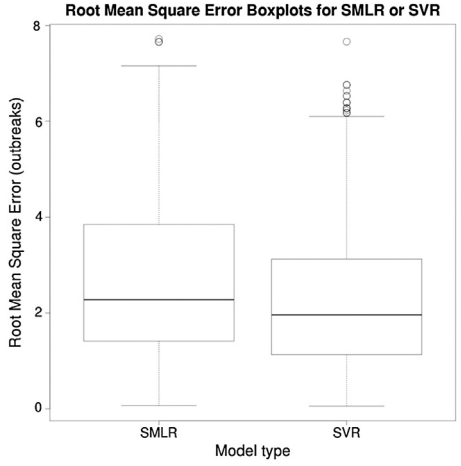

Figure 8. Box and whisker plots of RMSE bootstrap replicates for SMLR (on the left) and SVR (on the right). The SVR model displays a slightly lower median RMSE value and better consistency among model predictions.

outbreaks improvement in RMSE, but the more important finding was the smaller spread in the RMSE bootstrap replicates (IQR of 1.99). This smaller spread suggested greater consistency in the SVR predictions, which is important in such predictions. The RMSE was less platykurtotic (−0.08), though it was also skewed towards higher values (0.88). The peakedness near the median is ideal in a modeling study as it supports the other conclusions made that the SVR is yielding results consistent with the median RMSE. Note that R^2 statistics were not formulated for the SVR, since the R^2 is a linear statistic and SVR is a nonlinear regression method, making its interpretation difficult. Overall, it is clear that the introduction of nonlinearity and a machine learning method has improved both predictability and consistency of predictions in tornado outbreak forecasts at a one-year lag.

The bootstrap median residuals of the SVR and SMLR were computed as well (Fig. 9), following the same sampling method as the RMSE data. Overall, the SVR seemed to consistently slightly under predict outbreaks (median value of 0.133 outbreaks), while the SMLR model was consistently near zero (median value of 0.018 outbreaks). However, the SVR model had a smaller bootstrap median residual IQR (2.43) relative to the SMLR (2.91), suggesting that with small bias corrections, the SVR model could be improved further (this will be considered in future work).

In addition to the overall RMSE results, results by year were formulated by removing the year of interest from the training set, training on the remaining years, and then testing on the year of interest. This approach yielded individual predictions for each of the 51 years from both the SVR and SMLR models. Time series of the observed outbreak frequencies and the predicted frequencies for both the SVR (Fig. 10) and the SMLR (Fig. 11) revealed consistencies between the two modeling techniques regarding which years were predicted better or worse. Generally speaking, SMLR provided greater variability in predictions than SVR, which likely translated to its lower RMSE values. Additionally, both methods revealed the same upward trend present in the observed outbreak frequency time series (0.75 outbreaks per year), though the SMLR characterized this trend much more effectively than the SVR (0.69 outbreaks per year for SMLR, 0.52 outbreaks per year for SVR). Regardless, the smaller spread in the RMSE results and overall smaller RMSE values suggest the SVR is the better method to use, at least for short term predictions that are not sensitive to inconsistencies in the longer term trend.

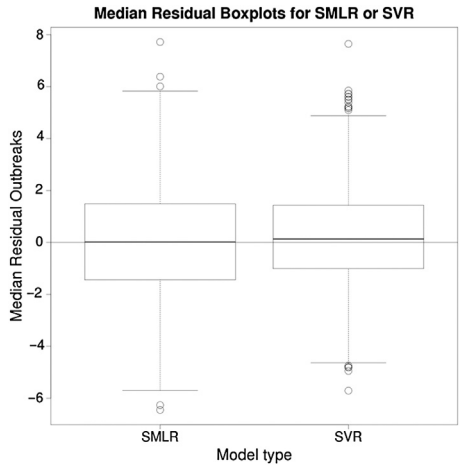

Figure 9. Box and whisker plots displaying the median residuals of the SMLR (left) and SVR (right).

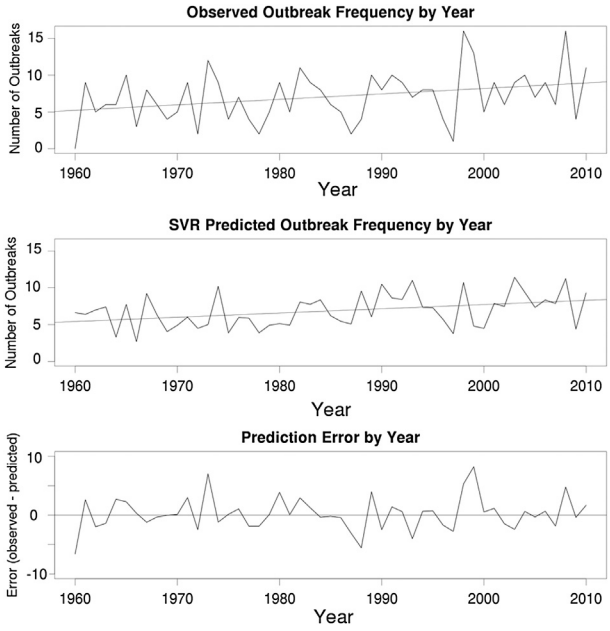

Figure 10. Time series of SVR performance based on the observed outbreak frequency (top panel), SVR predicted outbreak frequency (middle panel), and the prediction error (observed − predicted) for the SVR (lower panel). All SVR predictions are based on instances where the given year was withheld as independent testing and the remaining years were used as training.

3.2. Individual contributions by predictors and physical reasoning

To establish the relative contribution of each of the predictors to total outbreak number in the SMLR, the number of years that each predictor contributed one or more outbreaks was determined (based on the SMLR model coefficients; Table 3), as well as the number of years that each predictor led to a reduction of the number of outbreaks by one or more. Of course, since this is a linear model, some predictors contributed positively at the same time that others contributed negatively. This approach simply identifies the relative contributions by each predictor individually to try to

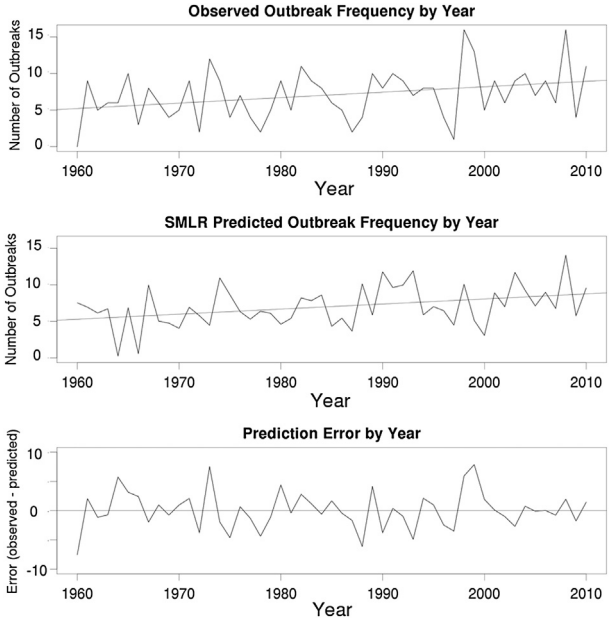

Figure 11. Same as Fig. 9, but for the SMLR model.

Table 3
Contributions of ±1 outbreak by predictor for the SMLR model, along with SMLR model coefficients. The + column represents the number of times 1 or more outbreaks was added to the overall prediction by the given variability index, while the − column represents the number of times 1 or more outbreaks was removed from the overall prediction.

Number of contributions of ±1 outbreaks by variability index			
Variability index	Coefficient	+1	−1
Jan. RPC 2	−0.7543	4	5
Jan. RPC 7	−1.3993	11	16
Feb. RPC 4	1.3518	12	13
Feb. RPC 8	−1.0233	10	7
Jan. RPC 3	−1.3369	9	11
Feb. AO	−0.2703	0	0
Feb. RPC 3	−1.5451	13	18
Jan. RPC 1	0.8405	2	6
Jan. ENSO	3.9068	22	19
Feb. ENSO	−4.1288	21	20
Feb. RPC 6	−0.4719	2	3

gain some physical insight into the model performance. Results are shown in Table 3. Note that a similar table for SVR cannot be formulated, since SVR arranges the predictors in a nonlinear manner.

The stepwise regression suggested that both ENSO patterns were significant, a result that supports other conclusions throughout the literature regarding the importance of ENSO phase on the position of the North American polar jet (Cook and Schaefer, 2008; Lee et al., 2013). The importance of ENSO phase was revealed by its SMLR model coefficient and its associated contributed/removed numbers from Table 3. Additionally, the February AO pattern was found to be an important predictor of tornado outbreaks, a result that is unsurprising given the AO's influence on Rossby wave patterns over North America (Thompson and Wallace 1998). However, the February AO index never contributed or removed 1 or more outbreaks from a given season, and thus is less important than some other predictors. The remaining 8 indices were based on the 500-hPa height fields, as described below.

January RPC1, RPC2, RPC3, and RPC7 all were selected as significant 500-hPa patterns for the month of January. This result is consistent with the areas of activity in each RPC, in particular RPC1, with a clear influence on North American Rossby wavetrains due to its representation of the strength of the Aleutian low, and RPC2, which has a direct center of activity over New England that is inversely correlated with a center of activity over the southern Plains. Interestingly, RPC1 contributed more negatively towards outbreak occurrence (6 years where it led to a reduction of outbreaks by 1 or more), while RPC2 had no such behavior. However, the individual contributions of both RPCs were minimal compared to other predictors (Table 3). RPC3 reveals little activity over North America, save for the small negatively correlated region over the Mid-Atlantic region which may be related to outbreaks in the eastern third of the US. However, this pattern contributed largely to increases during high years and decreases during low years (Table 3). RPC7's primary areas of activity are over the Eurasian continent as well, though a positive phasing of RPC7 contributes to negative height anomalies over the eastern US, creating northwest flow scenarios that are not ideal for tornado outbreaks. The importance of this pattern is evident, as it contributed to 1 or more outbreaks to 11 outbreak seasons and led to a decrease of 1 or more outbreaks in 16 seasons (Table 3).

The important February patterns (RPC3, RPC4, RPC6, and RPC8) reveal similar characteristics as those for January. RPC3, which closely matches the PNA pattern as defined by Richman and Mercer (2012), has been shown in the literature to have a direct impact on wave patterns over North America (Barnston and Livezey, 1987; Richman and Mercer, 2012). Similarly to January RPC1, this pattern contributed more negatively than positively to outbreak occurrence, but still contributed positively for 13 seasons (Table 3). Likewise, RPC6 reveals several areas of activity over North America, likely corresponding to the presence of an amplified wave train over the continent during that time that could lead to increased outbreak activity in the spring. However, this pattern's contributions and impacts were minimal. RPCs 4 and 8 show more indirect relationships, as RPC4 suggests modification of the Icelandic low, which could propagate upstream over North America, and RPC8 shows two weakly correlated areas off the Pacific and Atlantic coasts, suggesting the presence of a weak broad-scale trough that could be conducive for outbreak activity. RPC4 made large contributions to the outbreak forecasts as well, suggesting the outbreak forecasts are highly sensitive to this pattern (Table 3). However, RPC8 tended to contribute more positively (10 outbreaks) than negatively, and the SMLR was not as sensitive to its phase as RPC4.

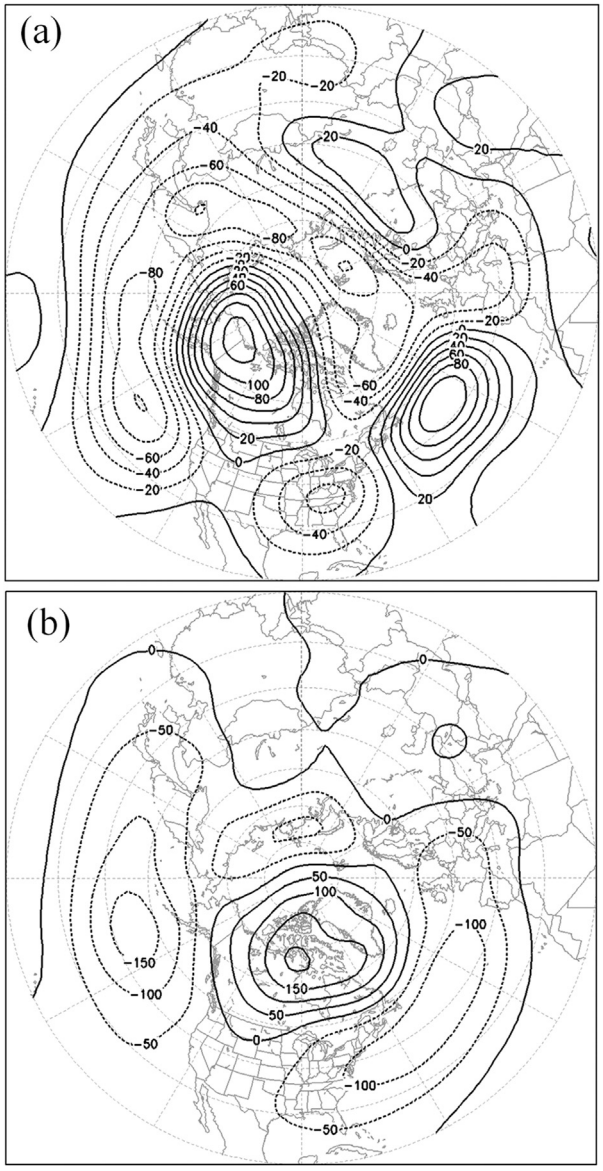

Figure 12. January (a) and February (b) 500-hPa unstandardized anomalies (calculated as annual mean−mean of all years) for 1990, a year prior to the highly active outbreak season of 1991. Negative contours suggest lower geopotential heights in January and February relative to the mean, likely supportive of an equatorward shift of the jet stream.

To further diagnose physical relationships between the indices and outbreak frequency, the January and February 500-hPa geopotential height fields for the year prior to a highly active outbreak season (1991) and a highly inactive outbreak season (1966) were retained. Unstandardized anomaly patterns of both the January and February fields (Fig. 12) reveal areas of generally low geopotential heights over most of North America, which suggests an equatorward shift of the jet stream. In addition, the subsequent year's May (May 1991; Fig. 13), which is the peak month for outbreak activity, revealed a strong negative anomaly over the central United States, providing an environment supportive of outbreaks consistent with Mercer et al. (2012). Similar results were found for other highly active outbreak years that were well predicted by the SVR and SMLR (e.g. 1982, 1998, 2010).

As expected, the results for the selected highly inactive year (1966) were consistent with an overall weaker outbreak season. The January and February 1965 geopotential height fields (Fig. 14) both revealed areas of above-average geopotential heights, suggesting that the jet stream during these months was shifted poleward, creating a more subtropical environment less suitable for outbreaks in the following months. These results were consistent with those observed for May of 1966 (Fig. 15), which shows a deep trough over the far eastern US, embedding the primary areas for outbreak activity in northwest flow, which is not conducive for outbreak formation. Similar results were seen with other well-predicted inactive years (e.g. 1962, 1977, 1979), revealing either a poleward shift in the jet stream or a longitudinally displaced Rossby wavetrain that is unsupportive of outbreak occurrence over the United States. Despite these examples, it is important to note that any physical relationships are speculative, and more work will need to be done to establish the physical mechanisms governing each pattern more robustly.

3.3. Independent testing years

To demonstrate model performance on independent testing years (2011–2013), the geopotential height indices were mapped onto the RPC loading vectors of the data trained from 1960 to 2010, and Niño 3.4 indices for these years were retained. The SMLR and SVR models trained on 1960–2010 were used in predicting 2011–2013 with these predictors. Unfortunately, 2011 was forecast poorly by both models, since 2011 was by far the most active outbreak season of the full 54 years of record. However, 2012 and 2013 were forecast well, with only small errors on the order of 2 outbreaks per year for each season (Fig. 16). These results suggest direct application of the results in predicting the onset of future outbreaks, which provide guidance not previously available to long-range forecasters. Additionally, since SVR is a learning method, incorporating years such as 2011 into the training database should help improve the forecasts further as more years are obtained.

Despite the seemingly strong relationships between the geopotential height indices and the Rossby wavetrain in subsequent months, the results of this study do not directly link the two from a physical standpoint. Likewise, these models do not predict any individual outbreak, but instead reveal the importance of the phasing of the Rossby wavetrain as a predictor for an upcoming

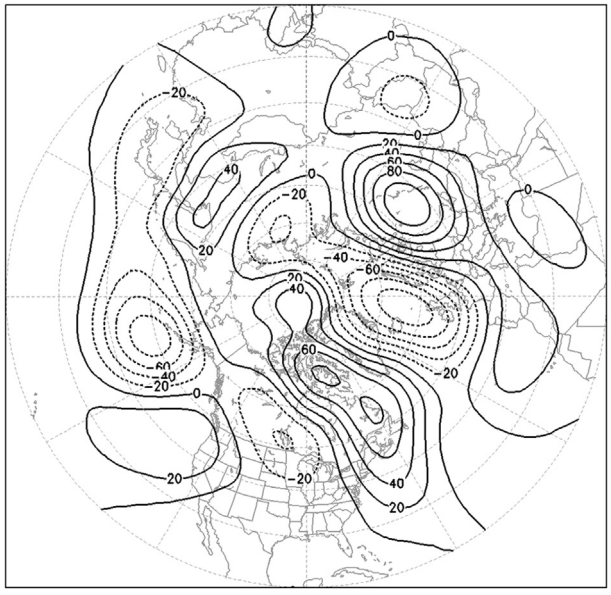

Figure 13. 500-hPa unstandardized anomalies for May 1991, the peak of the highly active outbreak season. Note the large area of negative anomalies over the United States, suggesting an amplified Rossby wavetrain supportive of increased outbreak frequency.

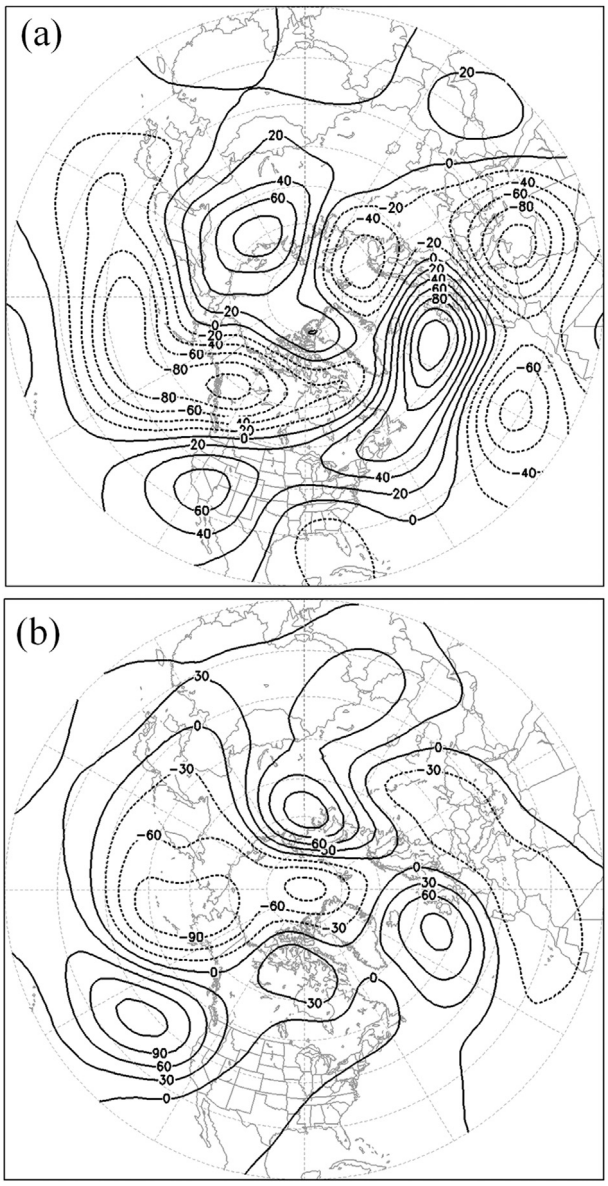

Figure 14. Same as Fig. 12, but for the year prior to a highly inactive season (1965).

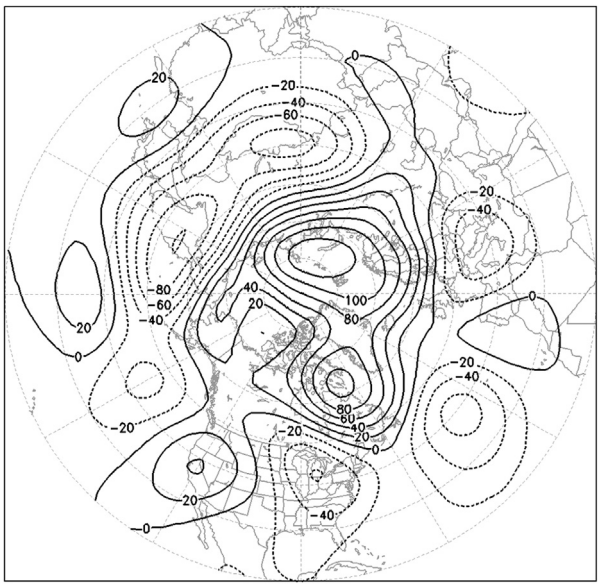

Figure 15. Same as Fig. 13, but for a highly inactive outbreak season (1966).

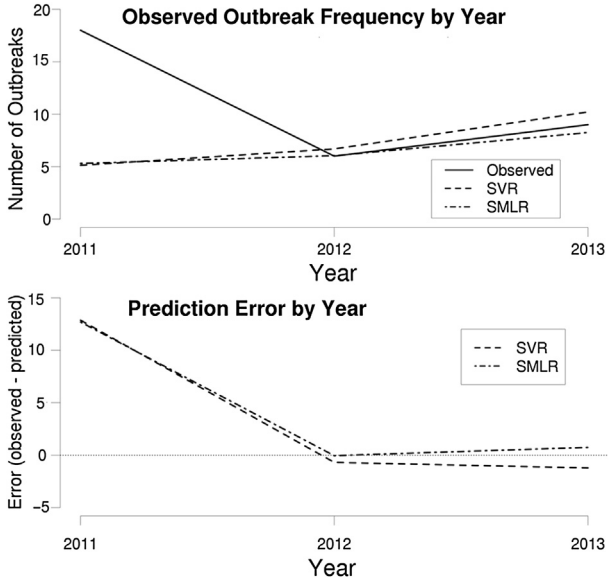

Figure 16. The top panel shows observed outbreak occurrence (solid line) against forecasted numbers of outbreaks for SVR (dashed line) and SMLR (dot-dash line) for independent test years 2011–2013. The bottom panel shows the forecast error for SVR (dashed line) and SMLR (dot-dashed line).

outbreak season with modest success. Future work will consider the detailed physical relationships between the individual patterns and outbreak frequency, but such ideas are outside the scope of this modeling study.

4. Summary/Conclusions

The purpose of this study was to diagnose seasonal predictability of tornado outbreak frequency in the United States through 500-hPa, 1000-hPa, and ENSO interannual variability indices. To diagnose the predictability of tornado outbreak frequency by year, an SMLR and an SVR model were formulated using 16 500-hPa RPC score predictors (8 from January and 8 from February), two 1000-hPa RPC score predictors (the AO pattern from January and February), and the Niño 3.4 indices for January and February. A forward-selection stepwise procedure was used to establish the best predictors, yielding a subset of 11 from the original 20. Upon determining the optimal predictor set, both SVR and SMLR models were trained to forecast seasonal outbreak frequency. Results revealed that the SVR model had lower overall RMSE, smaller spread in RMSE values, and the high model accuracy at a one year lag seems to suggest that the influence of teleconnections on planetary wavetrain patterns carries well over into the next year/outbreak season.

This study had several limitations associated with it. Shafer and Doswell (2010) indicated that the anthropogenic trend in tornado reports was minimized within their ranking scheme by applying specific weights to each variable. While the anthropogenic trend was minimized, there is still a statistically significant positive trend in outbreak frequency. It is likely that the trend can be attributed to the remaining anthropogenic component that cannot fully be removed, as well as climate variability that has resulted in more outbreaks in the past few decades. Additionally, while the physical links between outbreaks and the geopotential wavetrain seem to be present in the results, this study is only able to provide speculative insight as to the physical mechanisms behind the significant predictors in this study. Formally establishing these relationships is outside the scope of this study and will be considered in future work. Regardless, the results clearly demonstrate considerable skill in predicting the upcoming tornado outbreak year's outbreak frequency, an outcome that is useful to both forecasters and the general public in terms of preparing for the upcoming severe weather season. The hope is that the resulting models will be useful in implementing current operational seasonal outbreak forecasts.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. DGE-0947419 at Mississippi State University. We sincerely thank Dr. Chad Shafer for providing an updated N15 index. This work was improved greatly by the comments and suggestions from Dr. Grady Dixon and Dr. Jamie Dyer. We also would like to thank the reviewers for their thoughtful comments and suggestions that helped improve this work immensely.

References

Barnes, S., 1964. A technique for maximizing details in numerical weather-map analysis. Journal of Applied Meteorology 3, 396–409.
Barnston, A., Livezey, R., 1987. Classification, seasonality and persistence of low frequency atmospheric circulation patterns. Monthly Weather Review 115, 1083–1112.
Billet, J., DeLisi, M., Smith, B., Gates, C., 1997. Use of regression techniques to predict hail size and the probability of large hail. Weather and Forecasting 12, 154–164.
Colquhoun, J., Riley, P., 1996. Relationships between tornado intensity and various wind and thermodynamic variables. Weather and Forecasting 11, 360–371.
Cook, A.R., Schaefer, J.T., 2008. The relation of El Niño-Southern Oscillation (ENSO) to winter tornado outbreaks. Monthly Weather Review 136, 3121–3137.
Cristianini, N., Shawe-Taylor, J., 2000. An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods. Cambridge University Press, Cambridge, England, p. 189.
Dixon, P., Mercer, A., Grala, K., Cooke, W., 2014. Objective identification of tornado seasons and ideal spatial smoothing radii. Earth Interactions 18, 1–15.
Doswell, C.A., Edwards, R., Thompson, R.L., Hart, J.A., Crosbie, K.C., 2006. A simple and flexible method for ranking severe weather events. Weather & Forecasting 21, 939–951.
Durkee, J.D., Frye, J.D., Fuhrmann, C.M., Lacke, M.C., Jeong, H.G., Mote, T.L., 2008. Effects of the North Atlantic Oscillation on precipitation-type frequency and distribution in the eastern United States. Theoretical and Applied Climatology 94, 51–65.
Hagemeyer, B.C., 2011. The 2009-2010 El Niño and Florida Dry Season Tornadoes: A Reality Check for the Limits of Predictability. Available online at: http://www.srh.noaa.gov/media/mlb/pdfs/35cdpw-bhagemeyer.pdf.
Kalnay, E., Coauthors, 1996. The NCEP/NCAR 40-Year reanalysis project. Bulletin of the American Meteorological Society 77, 437–471.

Lee, S.K., Atlas, R., Enfield, D., Wang, C., Liu, H., 2013. Is there an optimal ENSO pattern that enhances large-scale atmospheric processes conducive to tornado outbreaks in the US? Journal of Climate 26, 1626–1642.

Linkin, M.E., Nigam, S., 2008. The north pacific oscillation-west Pacific teleconnection pattern: mature-phase structure and winter impacts. Journal of Climate 21, 1979–1997.

Mercer, A.E., Richman, M.B., Bluestein, H.B., Brown, J.M., 2008. Statistical modeling of downslope windstorms in Boulder, Colorado. Weather & Forecasting 23, 1176–1194.

Mercer, A.E., Shafer, C.M., Doswell III, C.A., Leslie, L.M., Richman, M.B., 2009. Objective classification of tornadic and nontornadic severe weather outbreaks. Monthly Weather Review 137, 4355–4368.

Mercer, A.E., Shafer, C.M., Doswell III, C.A., Leslie, L.M., Richman, M.B., 2012. Synoptic composites of tornadic and nontornadic outbreaks. Monthly Weather Review 140, 2590–2608.

Ohring, S., 1972. Application of stepwise multiple regression techniques to inversion of Nimbus "IRIS" observations. Monthly Weather Review 100, 336–344.

Richman, M.B., 1986. Rotation of principal components. Journal of Climate 6, 293–335.

Richman, M.B., Lamb, P.J., 1985. Climatic pattern analysis of three- and seven-day summer rainfall in the central United States: some methodological considerations and a regionalization. Journal of Climate and Applied Meteorology 24, 1325–1343.

Richman, M.B., Mercer, A.E., 2012. Identification of Intraseasonal Modes of Variability Using Rotated Principal Components, Atmospheric Model Applications. In: Dr. Ismail Yucel (Ed.), ISBN: 978-953-51-0488-9,InTech. Available from: http://www.intechopen.com/books/atmospheric-model-applications/identification-ofintrasesonal-modes-of-variability-using-rotated-principal-components.

Shabbar, A., Huang, J., Higuchi, K., 2001. The relationship between the wintertime North Atlantic Oscillation and blocking episodes in the North Atlantic. International Journal of Climatology 21, 355–369.

Shafer, C.M., Doswell, C.A., 2010. A multivariate index for ranking and classifying severe weather outbreaks. E-Journal of Severe Storms Meteorology 5, 1–39.

Smith, S.R., Green, P.M., Leonardi, A.P., O'Brien, J.J., 1998. Role of multiple-level tropospheric circulations in forcing ENSO winter precipitation anomalies. Monthly Weather Review 126, 3102–3116.

Swinbank, R., Purser, R.J., 2006. Fibonacci grids: a novel approach to global modelling. Quarterly Journal of the Royal Meteorological Society 132, 1769–1793.

Thompson, R.L., Vescio, M.D., 1998. The Destruction Potential Index – a Method for Comparing Tornado Days. Preprints - 19th Conference on Severe Local Storms. American Meteorological Society, Minneapolis, MN, pp. 280–282.

Thompson, D.W., Wallace, J.M., 1998. The Arctic Oscillation signature in the wintertime geopotential height and temperature fields. Geophysical Research Letters 25, 1297–1300.

Wallace, J., Gutzler, D., 1981. Teleconnections in the geopotential height field during the Northern Hemisphere winter. Monthly Weather Review 109, 784–812.

Wilks, D., 2011, third ed.. Statistical Methods in the Atmospheric Sciences Burlington, MA, vol. 100, p. 704.